

University of Groningen

## Responsiveness of the Roland-Morris Disability Questionnaire

Kuijer, W; Brouwer, S; Dijkstra, PU; Jorritsma, W; Groothoff, JW; Geertzen, JHB

*Published in:*  
Clinical Rehabilitation

*DOI:*  
[10.1191/0269215505cr842oa](https://doi.org/10.1191/0269215505cr842oa)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Early version, also known as pre-print

*Publication date:*  
2005

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Kuijer, W., Brouwer, S., Dijkstra, PU., Jorritsma, W., Groothoff, JW., & Geertzen, JHB. (2005). Responsiveness of the Roland-Morris Disability Questionnaire: consequences of using different external criteria. *Clinical Rehabilitation*, 19(5), 488-495. <https://doi.org/10.1191/0269215505cr842oa>

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# Clinical Rehabilitation

<http://cre.sagepub.com/>

---

## **Responsiveness of the Roland–Morris Disability Questionnaire: consequences of using different external criteria**

W Kuijer, S Brouwer, P U Dijkstra, W Jorritsma, J W Groothoff and J HB Geertzen

*Clin Rehabil* 2005 19: 488

DOI: 10.1191/0269215505cr842oa

The online version of this article can be found at:

<http://cre.sagepub.com/content/19/5/488>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *Clinical Rehabilitation* can be found at:**

**Email Alerts:** <http://cre.sagepub.com/cgi/alerts>

**Subscriptions:** <http://cre.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://cre.sagepub.com/content/19/5/488.refs.html>

>> [Version of Record](#) - May 1, 2005

[What is This?](#)

# Responsiveness of the Roland–Morris Disability Questionnaire: consequences of using different external criteria

**W Kuijer, S Brouwer** Centre for Rehabilitation, University Medical Centre Groningen and Northern Centre for Health Care Research, University of Groningen, **PU Dijkstra** Centre for Rehabilitation, University Medical Centre Groningen, Northern Centre for Health Care Research, University of Groningen and Department of Oral and Maxillofacial Surgery, University Medical Centre Groningen, **W Jorritsma** Centre for Rehabilitation, University Medical Centre Groningen, **JW Groothoff** Northern Centre for Health Care Research, University of Groningen and **JHB Geertzen** Centre for Rehabilitation, University Medical Centre Groningen and Northern Centre for Health Care Research, University of Groningen, The Netherlands

Received 28th February 2004; returned for revisions 23rd June 2004; revised manuscript accepted 25th July 2004.

**Objective:** To determine the consequences of using different external criteria on responsiveness of the Roland–Morris Disability Questionnaire (RMDQ) in patients with chronic low back pain.

**Design:** Questionnaire measures before and after rehabilitation treatment.

**Setting:** Rehabilitation centre.

**Subjects:** Patients with nonspecific chronic low back pain, referred for treatment.

**Main measures:** The RMDQ was used to assess self-reported functional status. The used external criteria were: (1) global perceived effect of change in complaints; (2) global perceived effect of change in ability to take care of oneself; (3) change in rating of pain intensity; (4) smallest real difference. Standardized response means, pooled effect sizes and receiver operating curves were calculated to determine responsiveness and to enable comparison of effect sizes with the thresholds of Cohen.

**Results:** Standardized response means ranged from 1.33 to 3.45, pooled effect sizes ranged from 1.50 to 2.81, and areas under curves ranged from 0.76 to 1.00, dependent on the used external criterion.

**Conclusions:** All pooled effect sizes were well above 0.80, and all other statistics were high, indicating good responsiveness of the RMDQ. However, considerable differences were found in responsiveness, when using different external criteria in a same study population. Therefore, it can be concluded that the magnitude of the responsiveness statistic depends on the used external criteria.

## Introduction

The Roland–Morris Disability Questionnaire (RMDQ) is often used as an evaluative outcome measure in patients with chronic low back pain<sup>1–3</sup>

to assess change in self-reported functional status after treatment. Evaluative outcome measures should be reliable and responsive to be able to assess change.<sup>4–6</sup> Both the English and Dutch versions of the RMDQ-24 show good reliability when using a time interval < three weeks (Pearson's  $r = 0.83$  and intraclass correlation coefficients range from 0.79 to 0.91).<sup>7–12</sup> Despite the good reliability, smallest real differences of 5.4 and 5.9

Address for correspondence: W Kuijer, Centre for Rehabilitation, University Medical Centre Groningen, PO Box 30.001, 9700 RB, Groningen, The Netherlands.  
e-mail: w.kuijer@rev.umcg.nl

show that substantial variation must be taken into account when the RMDQ is used in a clinical setting.<sup>7,13</sup> The responsiveness of the English and Dutch versions of the RMDQ has also been investigated in several studies. However, different outcomes were found for both versions. Responsiveness statistics, such as areas under curves, ranged from 0.68 to 0.84 for the English version,<sup>14–17</sup> and from 0.68 to 0.93 for the Dutch version.<sup>1,13,18</sup> Effect sizes ranged from 0.50 to 1.60 for the English version<sup>14,19–21</sup> and from 0.58 to 2.02 for the Dutch version.<sup>18,22</sup>

No clarity exists in the literature about responsiveness.<sup>23,24</sup> Different external criteria are used to determine whether a patient has achieved a clinically important change in functional status due to the absence of a gold standard to measure clinically important change. Furthermore, the terminology and calculation of responsiveness statistics as effect sizes and standardized response means may vary considerably,<sup>12</sup> depending on the type of change that is intended to be measured,<sup>10,24</sup> change in general, clinically important change, or the ability to detect changes in the construct being measured.<sup>24</sup> Previous studies evaluating the responsiveness of the RMDQ, used different external criteria, such as a global rating scale of magnitude in or importance of change (15 points),<sup>17,21</sup> a global rating scale of change in low back pain (7 points)<sup>1,13</sup> and a global rating scale of change in complaints (5 points).<sup>18</sup> It is not known to what extent different external criteria influence responsiveness statistics. In addition, whether the previously found effect sizes are trivial ( $\leq 0.20$ ), small ( $\geq 0.20$ ,  $< 0.50$ ), moderate ( $\geq 0.50$ ,  $< 0.80$ ) or large ( $\geq 0.80$ ), is not clear, because not all studies used the pooled effect size, as is needed for a direct comparison with these thresholds of Cohen.<sup>8</sup>

Because the lack of clarity about the definition of responsiveness, different external criteria<sup>1,8,21,22,25</sup> and unambiguous responsiveness statistics should be used in the same study to determine the responsiveness of an evaluative outcome instrument. The aim of this study is to determine the consequences of using different external criteria on the responsiveness of the RMDQ in patients with chronic low back pain.

## Methods

### Subjects

An existing clinical database was used. Data were gathered before and after rehabilitation treatments of patients with nonspecific chronic low back pain at the Centre for Rehabilitation location Beatrixoord in Haren, the Netherlands. In total, 83 patients (44 men and 39 women) with a mean age of 38.5 years (SD 9.7) participated and filled out the RMDQ before and after treatment. The mean RMDQ score before treatment was 10.9 (SD 4.7). The mean duration of treatment was 28 weeks (SD 14.5), including one to two treatment sessions per week.

### Outcome measure

The Dutch version of the RMDQ-24 was used to assess self-reported functional status before and after treatment. The RMDQ is derived from the Sickness Impact Profile, a general health questionnaire,<sup>26</sup> and assesses, dichotomously, perceived limitations due to low back pain in 24 activities of daily living. The time frame used in this study was 'the past few days'. The sum score is calculated by summing the 'yes' answers. The scale ranges from zero (no disability) to 24 (severe disability).

### External criteria

Different external criteria were used to analyse the responsiveness of the RMDQ. First, the global perceived effect of change in complaints due to chronic low back pain (7-point scale, ranging from 'completely recovered' to 'worse than ever'). Patients were improved if they scored 'completely recovered' or 'much recovered'. Second, the global perceived effect of change in ability to take care of oneself (4-point scale, ranging from 'much improved' to 'not improved'). Patients were improved if they scored 'much improved'. Third, change in rating of pain intensity. Three 10-point pain intensity scales were used before and after treatment, representing 'pain when at worst', 'pain when at least' and 'pain right now'. All scales were used as a separate external criterion. Additionally, the mean of the three scales represented the fourth pain intensity criterion. According to patients' ratings of pain intensity, patients were improved if the change score was 2 units or more on the pain scales.<sup>4,9,27</sup> Finally, the smallest real

difference of the RMDQ was used as an external criterion.<sup>13</sup> Patients were improved if their change score exceeded the smallest real difference of the RMDQ (i.e., if they changed 6 points or more on the RMDQ).<sup>7,13</sup> The external criteria on pain and complaints are impairment based, and the criteria on ability to take care of oneself is participation based.

### Data analyses

To calculate the association between the different external criteria, Spearman rank correlation coefficients were calculated. To calculate the association between the different external criteria and RMDQ change scores, Pearson correlation coefficients were calculated. Missing data were excluded pair wise from analyses. The smallest real difference criterion is not compared with the other external criteria and with the RMDQ change score, because this criterion is only a cut-off score. No correlation coefficients were calculated between mean pain intensity score and other pain scores, because the mean score is computed from the other pain scores. Therefore, these scores are not independent of each other.

Means and standard deviations of the RMDQ before and after treatment, and mean difference and the standard deviation of the difference were calculated for the different groups of patients classified as improved by the different external criteria. Standardized response means were calculated as the ratio of the mean difference of the improved group and the standard deviation of this mean difference. The higher the standardized response mean, the better the responsiveness. Pooled effect sizes were calculated as the ratio of mean difference of the improved group and the pooled standard deviation of the improved group ( $SD_{\text{pooled improved}}$ ), in which  $SD_{\text{pooled improved}} = \sqrt{[(SD_{\text{before treatment}}^2 + SD_{\text{after treatment}}^2)/2]}$ . Effect sizes are large when exceeding 0.80.<sup>8</sup> In addition, receiver operating curves were calculated. The receiver operator curve is a graph of 'true positive' (sensitivity) versus 'false positive' (1-specificity) for each of several cut-off points in score change.<sup>23,28</sup> The area under the receiver operating curve can be interpreted as the probability of correctly discriminating between improved and nonimproved patients. This area theoretically ranges from 0.5 (no accuracy in discriminating

improved from nonimproved patients) to 1.0 (perfect accuracy).<sup>28</sup> The four different external criteria were used to discriminate between improved and nonimproved patients.

An RMDQ score defined as 'no limitations' can vary between 0 and 5 points, based upon the established smallest real difference of the RMDQ. Therefore, it was decided to perform above-mentioned analyses not only in the total group of patients classified as improved according to the different external criteria, but also in improved patients with an initial RMDQ score  $\geq 6$ . Only these patients are 'certain' of having a limitation in self-reported functional status due to chronic low back pain (initial RMDQ score  $>$  smallest real difference), and can show improvement according to the smallest real difference of the RMDQ.

Furthermore, the number of improved patients classified by the different external criteria was calculated. Finally, the relationship between pain and self-reported functional status has been investigated in a post hoc analysis. Pearson correlation coefficients were calculated between the different initial pain intensity scores and initial RMDQ scores.

### Results

All results are shown for both improved patients and improved patients with an initial RMDQ score  $\geq 6$ , classified by the different external criteria. All improved patients classified by the smallest real difference criterion have an initial RMDQ score  $\geq 6$ . Therefore, only one group of patients is shown according to this criterion.

Spearman rank correlation coefficients between the different external criteria and Pearson correlation coefficients between the different external criteria and the RMDQ change score range from 0.27 to 0.85 (Table 1). Valid percentages of the improved patients are presented in Table 2, as well as means, standard deviations of initial, post and RMDQ change scores, effect sizes (standardized response means and pooled effect sizes), and areas under curves with confidence intervals. Standardized response means range from 1.33 to 3.45. Pooled effect sizes range from 1.50 to 2.81. Areas under curve range from 0.76 to 1.00. Sixteen

**Table 1** Relationships between change in self-reported functional status and different external criteria\*

	$\Delta$ RMDQ <sup>a</sup>	Complaints <sup>b</sup>	Taking care of oneself <sup>c</sup>	$\Delta$ Pain least <sup>d</sup>	$\Delta$ Pain worst <sup>e</sup>
Complaints <sup>b</sup>	– 0.59				
Taking care of oneself <sup>c</sup>	– 0.51	0.78			
$\Delta$ Pain mean <sup>f</sup>	0.85	– 0.56	– 0.43		
$\Delta$ Pain least <sup>d</sup>	0.66	– 0.41	– 0.27		
$\Delta$ Pain worst <sup>e</sup>	0.80	– 0.50	– 0.38	0.76	
$\Delta$ Pain now <sup>g</sup>	0.80	– 0.60	– 0.47	0.72	0.77

\*All coefficients were significant at  $p \leq 0.001$ .

<sup>a</sup>Change in self-reported functional status, measured by the RMDQ.

<sup>b</sup>Global rating of change in complaints.

<sup>c</sup>Global rating of change in taking care of oneself.

<sup>f</sup>Mean pain intensity difference.

<sup>d</sup>Pain intensity difference 'pain when at least'.

<sup>e</sup>Pain intensity difference 'pain when at worst'.

<sup>g</sup>Pain intensity difference 'pain right now'.

patients are classified as improved if all seven external criteria are applied. Twenty patients were improved on six criteria, 26 patients were improved on five criteria, 32 patients were improved on four criteria, 35 patients were improved on three criteria, 41 patients were improved on two criteria, 50 patients were improved on one criteria and 14 patients were not improved on a single criterion.

Correlation coefficients between pain intensity scores and RMDQ change score ranged between 0.43 and 0.56 (all were significant at  $p < 0.001$ , Table 3). Pearson correlation coefficients between different change pain intensity scales and RMDQ change score ranged from 0.66 to 0.85 (Table 1).

## Discussion

The choice of the external criterion influences the size of the responsiveness statistic. In our study, comparing application of external criteria in the same study population, considerable differences in effect sizes were found. The differences in effect sizes amounted to 2.12 points. In addition, the differences found in areas under curves amounted to 0.24. Apart from these large differences, comparison of pooled effect sizes with the thresholds of Cohen showed that all effect sizes ranged well above 0.80. Furthermore, all areas under curves were above 0.75. These results indicate good responsiveness of the RMDQ, independent of the used external criterion.

In the present study, using the smallest real difference as an external criterion, yields the highest statistic for both the effect sizes and the size of the area under curve. It was expected that using this statistic as a criterion for change, almost all patients would be classified as improved correctly, because smallest real difference as a cut-off score for change is not a real external criterion, but it is based on the measurement properties of the instrument itself. The specificity to change is with the smallest real difference as the cut-off for change per definition, for a normal distribution, equal to 95%.<sup>13</sup> In our study however, the specificity to change was 100%, due to a skewed distribution of RMDQ change scores. This means that the estimation, if patients underwent a clinically important change, can be made with 100% accuracy in this group of patients. The next highest responsiveness statistic is found when using mean pain intensity difference as an external criterion, followed by the pain intensity scales 'pain right now', 'pain when at least' and 'pain when at worst'. We chose to use all four pain intensity scales as separate external criteria because we wanted to compare different external criteria in this responsiveness study. The mean pain intensity criterion was previously used in a study after responsiveness of the 100-mm visual analogue scale, specifically aimed at patients with chronic low back pain.<sup>6</sup> It should be noted, however, that the validity of the calculation of this mean pain intensity has not been investigated. The high responsiveness statistic



**Table 2** Characteristics and responsiveness of the RMDQ, applying different external criteria

Criterion	Global rating of change in complaints	Global rating of taking care of oneself	Pain intensity difference 'pain when at least'	Pain intensity difference 'pain when at worst'	Pain intensity difference 'pain right now'	Mean pain intensity difference	Smallest real difference
<b>All improved patients</b>							
Improved (n [valid%])	39 [48.8]	37 [46.8]	38 [46.9]	30 [37.0]	36 [43.9]	31 [38.3]	
Initial RMDQ score (mean [SD])	11.3 [5.0]	11.4 [5.1]	12.7 [4.1]	12.2 [4.3]	13.2 [3.6]	13.3 [3.2]	
Post RMDQ score (mean [SD])	4.0 [3.6]	4.5 [4.5]	4.6 [3.7]	4.2 [3.4]	4.6 [3.9]	4.3 [3.4]	
Mean change in RMDQ (mean [SD])	7.4 [4.7]	6.9 [5.2]	8.1 [3.8]	8.0 [4.3]	8.6 [3.6]	9.0 [3.4]	
SRM <sup>a</sup>	1.56	1.33	2.16	1.87	2.42	2.64	
ES <sup>b</sup>	1.68	1.50	2.09	2.07	2.29	2.69	
AUC [95% CI] <sup>c</sup>	0.82 [0.73 to 0.91]	0.76 [0.65 to 0.87]	0.92 [0.86 to 0.98]	0.84 [0.75 to 0.92]	0.94 [0.89 to 0.98]	0.93 [0.88 to 0.98]	
<b>Improved patients with an initial score <math>\geq 6</math></b>							
Improved (n [valid%])	35 [51.5]	33 [49.3]	35 [51.5]	27 [39.7]	35 [50.7]	31 [45.6]	35 [50]
Initial RMDQ score (mean [SD])	12.8 [3.2]	12.8 [3.1]	13.5 [3.3]	13.2 [3.3]	13.5 [3.3]	13.3 [3.2]	13.6 [3.2]
Post RMDQ score (mean [SD])	4.6 [4.0]	5.1 [4.5]	4.8 [3.7]	4.4 [3.5]	4.7 [3.9]	4.3 [3.4]	4.2 [3.5]
Mean change in RMDQ (mean [SD])	8.2 [4.1]	7.8 [4.8]	8.7 [3.3]	8.8 [3.8]	8.7 [3.6]	9.0 [3.4]	9.4 [2.7]
SRM <sup>a</sup>	1.98	1.62	2.63	2.34	2.45	2.64	3.45
ES <sup>b</sup>	2.26	1.98	2.45	2.60	2.39	2.69	2.81
AUC [95% CI] <sup>c</sup>	0.85 [0.75 to 0.94]	0.78 [0.66 to 0.89]	0.93 [0.87 to 1.00]	0.85 [0.76 to 0.94]	0.92 [0.86 to 0.98]	0.92 [0.85 to 0.98]	1.00 [1.00]

<sup>a</sup>Standardized response mean = mean difference improved group/SD difference improved group.<sup>b</sup>Pooled effect size = mean difference improved group/SD<sub>pooled improved group</sub>, in which SD<sub>pooled improved group</sub> = root mean square of the standard deviations before and after treatment of the improved group.<sup>c</sup>Area under curve [95% confidence interval].

**Table 3** Relationships between self-reported functional status and initial pain scores<sup>a</sup>

	Initial RMDQ score
Initial mean pain intensity	0.56
Initial pain intensity 'pain when at least'	0.43
Initial pain intensity 'pain when at worst'	0.44
Initial pain intensity 'pain right now'	0.55

<sup>a</sup>All correlation coefficients were significant at  $p \leq 0.001$ .

found for the RMDQ when using the criteria for change in pain intensity means that a change in self-reported pain intensity accompanies a considerable change in self-reported functional status, as Beurskens already suggested.<sup>22</sup> However, initial pain scores (impairment based) and self-reported functional status measures (limitation based) are not strongly related (correlation coefficients ranged from 0.27 to 0.69<sup>1,5,6,11,18,22,29,30</sup>).

A discrepancy exists between clinical assessment of patients with chronic low back pain, and scientific purposes of determining the responsiveness of the RMDQ in this group of patients. Looking at the smallest real difference of the RMDQ, in our study, 10–11% of the patients are not 'certain' of having a limitation in self-reported functional status due to chronic low back pain (initial RMDQ score < smallest real difference), and cannot show any improvement according to the smallest real difference of the RMDQ. However, these patients were treated for their low back complaints, and they did improve according to the external criteria global rating of complaints, global rating of taking care of oneself and self-reported change in pain intensity. This discrepancy can be explained as follows: first, patients' limitations concern domains other than functional status in daily life, which is measured by the RMDQ. Secondly, an improvement on one of the items of the RMDQ is that important for the patient, that he or she feels really improved, despite the absence of change on other items, the so-called patient priority. For these reasons, patients did not improve in self-reported functional status as measured by the RMDQ, but they may have reached their personal treatment goals. Thirdly, the global perceived effect measures are only used *after* treatment. Because patients usually underestimate their pretreatment state, their assessment of being

### Clinical messages

- The choice of the external criterion determines the accuracy in discriminating improved from nonimproved patients.
- The Roland–Morris Disability Questionnaire can be used to determine whether a patient has changed in self-reported functional status, independent of the used external criterion.

improved after treatment may therefore be biased,<sup>31</sup> resulting in an overestimation of improvement on the global ratings. The pain intensity criteria scales are used before and after treatment, but the scales 'pain when at worst' and 'pain when at least' also refer to a previous state. Because this discrepancy between clinical assessment and research purposes, we decided to dichotomize our presentation of results into two groups: all improved patients and the group of improved patients with an initial RMDQ score  $\geq 6$ . Lower responsiveness statistics are found for all improved patients compared to the improved group with an initial RMDQ score  $\geq 6$ , when using global rating of change in complaints, global rating of taking care of oneself, and change in pain intensity scores 'pain when at least', 'pain when at worst' and 'pain right now' as external criteria.

These differences in responsiveness statistics between both groups can be explained by baseline score variability. If baseline score variability decreases, which occurs when excluding patients with an initial score  $\leq 6$ , the responsiveness of a measurement increases.<sup>32</sup> When using the criteria of mean pain intensity or smallest real difference, exactly the same patients were classified as 'improved' for all patients as for the group with an initial score  $\geq 6$ , thus no change in baseline score variability occurs. In addition, the ability of the RMDQ to detect changes diminishes when small limitations exist in self-reported functional status, as is measured by the RMDQ.<sup>17,33</sup>

The study has some limitations. First, the sample size of the study is somewhat small, in relation to the number of statistical tests performed. However, all results indicate large differences in responsiveness. Therefore, our conclusion



that the use of different external criteria leads to differences in responsiveness statistics is rather robust. Second, it might be argued that the RMDQ measures limitations in activities of daily living on an interval scale, ranging from zero limitations on the questioned activities to 24 activity limitations. However, no evidence exist that each item should be weighted to calculate a sum score. Therefore, in this study, parametrical statistical techniques are used for analyses of the RMDQ. This is commonly undertaken when analysing the RMDQ.

Good responsiveness is found for the RMDQ. However, considerable differences were found in responsiveness statistics when using different external criteria in a same study population. Therefore, it can be concluded that the magnitude of responsiveness statistics depends on the external criteria used.

### Acknowledgements

We thank the clinicians and the co-workers of the Centre for Rehabilitation, location Beatrixoord in Haren, the Netherlands, for their co-operation in this study. The study is supported by grants from Zorgonderzoek Nederland (Zon Mw), and the foundation 'Beatrixoord Noord-Nederland'.

### References

- 1 Beurskens AJ, Vet de HC, Koke AJ, Heijden van der GJ, Knipschild PG. Measuring the functional status of patients with low back pain. *Spine* 1995; **20**: 1017–28.
- 2 Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 2000; **25**: 3100–103.
- 3 Roland M, Fairbank J. The Roland–Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine* 2000; **25**: 3115–24.
- 4 Farrar JT, Young JPJr, Moreaux La L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain* 2001; **94**: 149–58.
- 5 Gronblad M, Jarvinen EAO, Ruuskanen M, Hamalainen H, Kouri J. Relationship of subjective disability with pain intensity, pain duration, pain location, and work-related factors in nonoperated patients with chronic low back pain. *Clin J Pain* 1996; **12**: 194–200.
- 6 Hazard RG, Haugh LD, Green PA, Jones PL. Chronic low back pain. The relationship between patient satisfaction and pain, impairment, and disability outcomes. *Spine* 1994; **19**: 881–87.
- 7 Brouwer S, Kuijer W, Dijkstra PU, Goeken LNH, Groothoff JW, Geertzen JHB. Reliability and stability of the Roland Morris Disability Questionnaire: intra class correlation and limits of agreement. *Disabil Rehabil* 2004; **26**: 162–65.
- 8 Cohen, J. *Statistical power analysis for the behavioral sciences*. New York: Academic Press, 1988: 1–27.
- 9 Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL. Defining the clinically important difference in pain outcome measures. *Pain* 2000; **88**: 287–94.
- 10 Liang MH. Evaluating measurement responsiveness. *J Rheumatol* 1995; **22**: 1191–92.
- 11 Takeyachi Y, Konno S, Otani K *et al.* Correlation of low back pain with functional status, general health perception, social participation, subjective happiness, and patient satisfaction. *Spine* 2003; **28**: 1461–67.
- 12 Taylor SJ, Taylor AE, Foy MA, Fogg AJB. Responsiveness of common outcome measures for patients with low back pain. *Spine* 1999; **24**: 1805–12.
- 13 Vet de HC, Bouter LM, Bezemer PM. Reproducibility and responsiveness of evaluative outcome measures. *Int J Technol Assess Health Care* 2001; **17**: 479–87.
- 14 Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther* 2002; **82**: 8–24.
- 15 Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland–Morris Back Pain Questionnaire: part 2. *Phys Ther* 1998; **78**: 1197–207.
- 16 Stratford PW, Binkley J, Solomon P, Gill C, Finch E. Assessing change over time in patients with low back pain. *Phys Ther* 1994; **74**: 528–33.
- 17 Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland–Morris Back Pain Questionnaire: part 1. *Phys Ther* 1998; **78**: 1186–96.
- 18 Gommans IHB, Koes BW, Tulder van MW. Validiteit en responsiviteit Nederlandstalige Roland Disability Questionnaire. Vragenlijst naar functionele status bij patiënten met lage rugpijn. (Validity and Responsiveness of the Dutch version of the Roland Disability Questionnaire. A functional status questionnaire for patients with low back pain.) *Ned Tijds Fysiother* 1997; **107**: 28–33.

- 19 Kopec JA, Esdaile JM, Abrahamowicz M *et al.* The Quebec Back Pain Disability Scale. *Spine* 1995; **20**: 341–52.
- 20 Patrick DL, Deyo R, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. *Spine* 1995; **20**: 1899–909.
- 21 Stratford PW, Binkley J, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther* 1996; **76**: 1109–23.
- 22 Beurskens AJ, Vet de HC, Koke AJ. Responsiveness of functional status in low back pain. A comparison of different instruments. *Pain* 1996; **65**: 71–6.
- 23 Deyo R, Centor R. Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance. *J Chronic Disabil* 1986; **39**: 897–906.
- 24 Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PMM. On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. *Qual Life Res* 2003; **12**: 349–62.
- 25 Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *J Clin Epidemiol* 1997; **50**: 79–93.
- 26 Bergner M, Bobbitt RA, Carter WB, Gilson BS. Sickness Impact Profile: Development and final revision of health status measure. *Med Care* 1981; **19**: 787–805.
- 27 Hagg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 2003; **12**: 12–20.
- 28 Sackett, DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology, a basic science for clinical medicine*. Boston/Toronto/London: Little, Brown and Company, 1991.
- 29 Gronblad M, Hupli M, Wennerstrand P *et al.* Intercorrelation and test–retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. *Clin J Pain* 1993; **9**: 189–95.
- 30 Koho P, Aho S, Watson P, Hurri H. Assessment of chronic pain behaviour: reliability of the method and its relationship with perceived disability, physical impairment and function. *J Rehabil Med* 2001; **33**: 128–32.
- 31 Linton SJ, Melin L. The accuracy of remembering chronic pain. *Pain* 1982; **13**: 281–85.
- 32 Guyatt G, Walter S, Norman G. Measuring change over time: Assessing the usefulness of evaluative instruments. *J Chronic Disabil* 1987; **40**: 171–78.
- 33 Stratford PW, Binkley J, Solomon P, Finch E, Gill C, Moreland J. Defining the minimum level of detectable change for the Roland–Morris questionnaire. *Phys Ther* 1996; **76**: 359–65.